FAIR 原则视角下我国科学数据管理政策研究
——以国家科学数据中心和高校机构知识库为例

杨恒[1，2] 刘凤红*[1，2，3]

1. 中国科学院大学经济管理学院图书情报与档案管理系北京 100191
2. 中国科学院文献情报中心北京 100191
3. 南京大学信息管理学院南京 210023

摘要：[目的/意义] 目前，FAIR 原则的理念已在一定程度上显性或隐性地蕴含在我国科学数据管理相关政策文件中，我国科学数据管理政策的 FAIR 化目前处在何种程度，在 FAIR 化过程中有何特点，未来应在何处继续着力等问题亟待探究。[方法/过程] 本文以国家科学数据中心和高校机构知识库发布的科学数据管理相关政策文件为样本，运用计量统计、文本挖掘等方法充分揭示其内外部特征，并提出一种基于"锚点词"的 FAIR 原则到政策文本对齐方法，对我国科学数据管理政策的整体 FAIR 化程度与特征做出研判。[结果/结论] 发现我国科学数据管理政策对 FAIR 原则的总体匹配程度偏低、对 FAIR 各原则的关注度呈三个层次的阶梯状分布、整体表现出一种"不均衡"的状态。

关键词：科学数据管理；FAIR 原则；文本对齐

1 引言

随着开放科学与开放获取运动[1]的兴起与发展，科学研究加速向数据驱动型的第四范式转变，科学数据已经成为最有价值的战略资源之一，是科技创新的关键要素并会带来社会效益[2]。相应的，科学数据管理与共享的 FAIR 原则[3]也已成为被国际社会广泛接受的准则之一。国内对 FAIR 原则的引入与研究方兴未艾，《中国科学院科学数据管理与开放共享办法（试行）》指出，"科学数据应按照分等级、可发现、可访问、可重用的原则，适时向院内外用户开放共享。[4]"《新一代人工智能伦理规范》也要求"提升数据的完整性、及时性、一致性、规范性和准确性等。[5]"可见，FAIR 的理念已在一定程度上显性或隐性地蕴含在我国相关政策文件中对科学数据的管理规范上。因此，我国科学数据管理政策的 FAIR 化目前处在何种程度，在 FAIR 化过程中有何特点，未来应在何处继续着力以对接 FAIR 原则等问题，都是在整个科研数据生态体系中,建立方便、可靠、可测量、可验证、可支持 FAIR 原则的数据出版机制与数据共享政策[6]所亟待澄清的。

基于此，本文以国家科学数据中心（以下简称"数据中心"）和高校机构知识库（以下简称"机构知识库"）发布的科学数据管理相关政策文件为样本，在充分了解其内外部特征的基础上，将其置于 FAIR 原则理论框架下做对齐分析，以求

对我国科学数据管理政策的整体 FAIR 化程度与特征做出研判，更好地保障科学数据在科研活动过程中发挥纽带和基石的作用[7]。

2 相关研究回顾与简评

学界关于“科学数据”的概念目前并未形成统一意见，从不同的定义基础[8]和相关性[9]出发，形成了多样化的命名方式，如：“科学数据（Scientific data）”“科研/研究数据（Research data）”“科技数据（Scientific and technological data）”等。本文选取的研究对象包括数据中心与机构知识库发布的数据政策文件，两者均有覆盖范围广、使用类型多样等特点，因此我们采用在概念内涵上较为宏观的“科学数据”对其进行指代，但对同行学者的参考仍保留其原表述。

本文选用中国知网作为检索工具，在“信息科技”类目下，分别用“科学数据管理政策“科研数据管理政策”“研究数据管理政策”以及“科技数据管理政策”为检索词进行主题检索（检索时间：2021 年 8 月 24 日），得到文献 110 篇。通过逐篇阅读的方式剔除不涉及科学数据管理政策的研究以及部分质量不高的文献，最终确定了 95 篇相关文献。从国内已有的论文成果来看，对科学数据管理政策的研究通常与数据权益[10][11]、数据素养、[12][13]数据共享[14][15][16]等术语挂钩，学者们从不同的角度切入，对国内外各类机构发布的科学数据管理政策进行了较为深入的剖析和研究，相关研究成果按照研究方法的不同，主要有三种类型：

（1）政策解读类。在对国内科学数据管理政策的研究中，以《科学数据管理办法》为样本对其进行解读是一大热点，邢文明[17][18][19]、高瑜蔚、[20]白锐[21]等对政策文本内容进行了详细解读，温亮明[22]、李洋[23]等对其落实路径与现状进行了分析与探讨。同时，对国外相关政策的介绍、梳理与分析也是十分活跃，且不仅仅局限于对国外政府发布的科学数据管理政策的研究，而是从教育机构[24]、科研资助机构[25]、科学联盟[26]等不同的角度展开探索，为我国科学数据管理政策的完善建言献策，如邢文明等[27]以《科学数据管理办法》和《促进联邦资助科研成果获取的备忘录》为例对中美科学数据政策进行了比较研究，提出建议引入科学数据开放的 FAIR 原则，以确保科学数据的可用性。

（2）调查实证类。我国的科学数据管理政策应与国际接轨，对国内外各类型利益相关者进行的调查实证类研究正是迎合了这样一种需求，因而该类型的研究活动十分活跃。然而，从调查主体的角度对比来看，国内作为调查对象的科研主体则略显单调，大部分的研究集中在高等院校[28][29]、高校图书馆[30]、机构

知识库[31]等科研机构领域；而对国外科学数据管理相关主体的调查，则更加涵盖了科研资助机构[32]、政府部门[33][34]、出版集团[35]、学术期刊[36][37]等各行各业的利益相关者。由此可见，仅就利益相关主体而言，国外的科学数据管理具有更为广泛的现实基础与配套设施，这更提醒我们无论是在科学数据管理的政策制定还是内容管理方面，都要时刻注重其宏观设计，将科学数据管理活动作为一个多方参与的整体进行规划与研究。

（3）专题研究类。科学数据管理政策方案的出台是对科研领域数据活动需求的应答，因此，除了“自上而下”地提出纲领设计之外，更重要的是对各领域的科学数据管理需求进行针对性的专题研究，“自下而上”地提出适用于实际需求的解决方案。如李芳薇等关注生物医学领域的科研数据仓储[38]与科学数据管理服务[39]；王卷乐[40][41]等人结合对全球 11 个方面的科学数据管理态势的分析，对我国地球科学领域的科学数据管理建设进行了展望；此外，陈廉芳认为需要建立科学数据馆员制度来规范和促进科学数据管理的发展[42]，并对医学领域的数据馆员制度建设进行了构想。学者们结合调查实证的研究方法，对各个科研细分领域的科学数据管理现状与需求进行了深入细致的分析，提出的意见与方案往往实用价值较高，因此这应该是未来科学数据管理研究的一大热点方向。

综上所述，学者们已经综合运用多种研究方法，从不同视角对我国科学数据管理政策进行了研究，产出了丰硕的学术成果。但是纵观我国学者对国内外科学数据管理政策的研究，往往由于缺乏相应的、完善的、来源于实践的理论作为指导与借鉴，从而难以形成契合实际的、接受度高的、泛化效果好的普适性方案。文献调研表明，面向科学数据管理活动整体的研究已有一定基础，如周玉琴等调查分析了国内外科研数据管理与共享政策体系建设的理论和实践现状，构建了一个由宏观层面的法律法规、中观层面的政策规章、微观层面的管理制度构成的科研数据管理与共享政策体系[43]；薛秋红等通过提炼西方国家科学数据管理政策的核心要素，试图构建起科学数据管理政策制定的基本规范[44]。上述研究为本文在 FAIR 数据管理原则的视角下，从理论整体出发，审视和考察我国科学数据管理相关政策，挖掘和梳理科学数据管理核心内容，探测和提出我国科学数据管理政策调整策略提供了有益借鉴。

3 研究设计

3.1 数据来源

2019 年 6 月，科技部、财政部公布国家科技资源共享服务平台优化调整名

单，确定自然科学与工程技术领域的 20 个国家科学数据中心[45]，这些数据中心主要依托科研院所管理运行，组成了我国科学数据管理与服务的“国家队”。2016 年 9 月 22 日，由 CALIS 管理中心联合 16 家高校图书馆共同发起成立了中国高校机构知识库联盟（CHAIR），以“推进全国高校机构知识库的建设，推动学术成果的开放获取，促进学术成果的广泛应用”为宗旨[46]，该联盟目前已注册有 51 家会员机构，元数据总量 2,868,428 条。本文以网络调研的方式对 20 家数据中心以及 51 家机构知识库联盟成员网站展开调查，围绕“科学数据管理”主题对其发布的数据政策进行爬梳剔抉，最终共获得网站公开发布的数据政策（全文）102 部，其中 90 部来自数据中心，12 部来自机构知识库，调查时间为 2021 年 9 月 10 日至 18 日。

本文选取数据中心和机构知识库的数据政策作为研究样本，主要基于两方面的原因：一是现阶段我国科学数据服务是以研究院所为主、高等院校及其图书馆为辅[47]，因此以二者为依托的数据平台的数据政策具有较强的代表性，对其进行研究有助于我们把握我国科学数据管理政策的整体发展态势，同时便于进行对数据政策内部的细粒度分析；另一方面，数据中心和机构知识库的数据政策分别代表了来自科研主体和教育主体对科学数据管理的理解与需求，在第四科研范式下二者既有交集也有差集，对其内容进行深度挖掘与分析可以使我们更好地认识到二者的相同点与差异性，从而为促进不同科学数据管理主体之间的协同发展、责任共担提供借鉴。

3.2 研究方法

本文围绕获取到的科学数据管理政策文本展开了两个方面的实证研究，一方面是数据中心和机构知识库政策文本的对比研究，具体分为三个步骤。1）对政策外部特征进行计量统计与分析，包括文件获取情况、发布时间等，主要用到了 Excel、ECharts 等工具；2）为了对政策文本的内部特征进行分析，对数据进行预处理。将从 95 篇“科学数据管理政策”相关文献中去重后得到的 194 个关键词作为保留词，百度停用词表作为停用词，建立起用于对切词内容进行清洗的用户词典；3）运用 Python-jieba、微思词云等工具，扫描政策文本中的高频词汇，绘制高频词云图，以分析对比数据中心与机构知识库数据政策的关注焦点，并初步探测其 FAIR 特征。

另一方面是 FAIR 原则视角下我国科学数据管理政策的对齐分析。Wilkinson 等为代表的“Go FAIR 指标小组”(Go FAIR Metric Group)从 14 个方面制定了

覆盖 15 条 FAIR 原则的评估量表，但是该量表的运用，要求来自关于被评估团体的多种行动信息[48]作为支撑，不适用于本研究对多来源大规模文本进行内容评估的需求。因此，本文提出一种基于“锚点词[49]”的 FAIR 原则到政策文本对齐方法，步骤如下：1）对不同版本的 FAIR 原则译文进行爬梳剔抉，根据同义关系、相关关系等为每条 FAIR 原则赋予或从中析出 FAIR 锚点词；2）采用字符串匹配的方式，用锚点词定位到政策文本中的相关句子，进行语句级别的文本对齐；3）对锚定获得的语句进行语义筛查，去除形同义不同的“伪锚点”，深入到语义级别的文本对齐；4）最后将对齐结果按机构进行统计整理，并进行可视化表示与分析讨论。

4 研究结果与讨论

4.1 我国科学数据管理政策内外部特征统计与分析

4.1.1 文件获取情况分析

由调查结果可知（见表 1 和表 2）：1）数据中心重视科学数据管理政策的建设，90%的科学数据中心均自行制定或援引了相关数据管理标准或政策，且能够使用户较为便捷地获取到（国家对地观测科学数据中心和国家材料腐蚀与防护科学数据中心除外），最终共从 18 家数据中心网站获取到了 90 部相关政策文件，其中 75 部（83.3%）能获取全文。2）各数据中心对科学数据管理政策建设的重视程度并不一致，有的出台了系列化、规范化的数据政策来支持科学数据管理活动全过程的顺利展开，如国家基因组科学数据中心、国家农业科学数据中心等；有的则只针对用户提交数据流程做了指导与说明，如国家空间科学数据中心、国家青藏高原科学数据中心等。3）机构知识库数据管理政策的建设现状不容乐观，其主要原因在于机构知识库本身建设状况不容乐观。尽管本文选取的调查样本为中国高校机构知识库联盟的 51 家成员机构，但仍有 33 家（64.7%）机构知识库的官方网站没有通过公开途径被访问到，其中 17 家“未找到网站”，16 家“网站无法访问”，在能够正常工作的机构知识库中，有 8 家（15.7%）没有获取到相关政策内容，其中 3 家“无权限访问”，5 家“无相关内容”，最后仅从 10 家（19.6%）机构知识库网站获取到了 12 部相关政策文件。

表 1 国家科学数据中心科学数据管理政策获取情况一览表

序号	平台名称	政策文件名	获取情况
1	国家高能物理	高能物理科学数据汇交管理办法	●

	科学数据中心	高能物理科学数据合作共享制度	●
2	国家基因组科学数据中心	原始组学数据标准	●
		基因组数据标准	●
		变异组数据标准	◑
		转录组数据标准	◑
		表观遗传数据标准	●
		异构数据索引和检索整合标准	●
3	国家微生物科学数据中心	国家微生物科学数据中心数据库访问协议	●
4	国家空间科学数据中心	空间科学数据汇交指南	●
5	国家天文科学数据中心	HDF5	◑
		普适图像传输系统（FITS）	◑
		Resource Metadata for the Virtual Observatory	●
		VOTable Format Definition	●
		IVOA 唯一标识符	●
		IVOA 锥形检索协议	●
		IVOA 简单图像访问协议	●
		IVOA 简单光谱访问协议	●
		IVOA 表格访问协议	●
		IVOA 目标可见性简单访问协议	●
		天文数据查询语言（ADQL）	●
		基于 HEALPix 的多层次天区覆盖图标准	●
		多层次渐进式巡天图	●
6	国家对地观测科学数据中心	-	○
7	国家极地科学数据中心	海洋观测规范 第 6 部分:数据处理与质量控制(修订稿)	◑
		水深测量数据采集与处理技术要求	◑
		海洋监测规范 第 2 部分:数据处理与分析质量控制	◑

		海洋声学数据共享交换格式	◑
		数据采集和处理规范	◑
		地理元数据信息	◑
		科学数据管理办法	●
		海洋信息元数据	◑
		声学多普勒流速剖面仪数据存储格式	◑
		海洋数据应用记录格式	◑
		中国极地考察数据管理办法	●
		国家极地科学数据中心元数据标准	●
8	国家青藏高原科学数据中心	汇交政策	●
		使用条款和免责申明	●
9	国家生态科学数据中心	网站介绍	●
10	国家材料腐蚀与防护科学数据中心	-	○
11	国家冰川冻土沙漠科学数据中心	中国特殊环境与灾害研究网络 数据管理与共享条例（征求意见稿）	●
		国家冰川冻土沙漠科学数据中心数据汇交协议	●
12	国家计量科学数据中心	国家计量科学数据中心项目数据汇交管理办法	●
		国家计量科学数据中心计量科研数据汇交标准规范	●
		国家计量科学数据中心数据分级分类管理办法	●
13	国家地球系统科学数据中心	数据汇交标准	●
14	国家人口健康科学数据中心	国家科技基础条件平台资源元数据核心元数据（征求意见稿）	●
		国家人口健康科学数据中心资源核心元数据（征求意见稿）	●
		国家人口健康科学数据中心科技资源标识符规范（征求意见稿）	●

		国家人口健康科学中心仓储元数据规范	●
		人口健康科学数据共享数据集分类与编码标准（征求意见稿）	●
		中华人民共和国数据安全法	●
		国家健康医疗大数据标准、安全和服务管理办法（试行）	●
		PHDA 数据资源收集保藏发展政策	●
		国家人口健康科学数据中心数据备份方案	●
15	国家基础学科公共科学数据中心	科学数据管理与服务标准规范	●
16	国家农业科学数据中心	农业科学数据采集标准编制要求	●
		农业科学数据发布管理规则	●
		农业科学数据中心数据服务规范	●
		农业科学数据质量检查与控制规范	●
		农业科学数据信息安全管理规范	●
		农业科学数据公共数据元标准	●
		农业科学数据汇交管理办法	●
		农业科学数据集成和访问规范	●
		农业科学数据交换格式规范	●
17	国家林业和草原科学数据中心	林业科学数据元数据标准（V3.10 版）	●
		林业科学数据集成规范（数据整合 XV2.0）	●
		林业科学数据分类与编码（V1.0）	●
		林业专题空间数据质量控制标准	●
		荒漠生态系统定位站观测数据规范	●
		林业科学数据数据字典规范	●
		林业科研机构基础数据规范	●
		科学数据共享工程质量管理规范	◑
		科学数据中心（网）运行管理规范	◑
18	国家气象科学数据中心	气象数据元 总则	●
		气象探测资料汇交管理办法	●

		汇交指南	●
19	国家地震科学数据中心	国家地震科学数据资源手册定稿（2020 版）	●
		地震科学数据共享管理办法	●
		地震科学数据　数据发布规范	●
		地震科学数据　共享系统运行规范	●
		地震科学数据　数据分类与编码	●
		数据科学数据　数据元目录	●
		地震科学数据　数据交换格式	●
		地震科学数据　数据模式编写	●
		地震科学数据　元数据编写指南	●
20	国家海洋科学数据中心	科技基础性工作专项项目科学数据汇交管理办法（试行）	●
		国家海洋科学数据共享服务平台建设运行管理暂行办法	●
		科技平台资源核心元数据（GBT 30523-2014）	●
		工业和信息化部关于工业大数据发展的指导意见	●
		涉海类科技计划项目科学数据汇交流程与规则说明	●
		中科院科学数据管理与开放共享办法	●

注：●表示有相关文件，且能获取全文；◑ 表示有相关文件，但无法获取全文；○表示无相关文件。

表 2 高校机构知识库科学数据管理政策获取情况一览表

序号	平台名称	政策文件名	获取情况	备注
1	北京大学机构知识库	北京大学机构知识库开放获取政策（试行）	●	
		北京大学机构知识库推荐文件格式	●	
		北京大学机构知识库提交流程	●	
2	北京化工大学机构	-	○	未找到

	知识库			网站
3	北京交通大学机构知识库	-	○	未找到网站
4	北京科技大学机构知识库	-	○	网站无法访问
5	北京理工大学机构知识库	-	◑	无相关内容
6	北京联合大学机构知识库	-	◑	无相关内容
7	北京师范大学学术成果库	北京师范大学学术成果库开放获取相关说明	●	
8	北京邮电大学机构知识库	-	○	网站无法访问
9	重庆大学机构知识库	-	○	未找到网站
10	大连理工大学机构知识库	大连理工大学机构知识库平台政策	●	
11	电子科技大学机构知识库	-	◑	无访问权限
12	东北师范大学机构知识库	-	○	未找到网站
13	东南大学机构知识库	-	◑	无访问权限
14	对外经济贸易大学机构知识库	-	○	未找到网站
15	福建工程学院机构知识库	-	○	未找到网站
16	哈尔滨工业大学机构知识库	-	◑	无相关内容
17	韩山师范学院机构知识库	-	○	未找到网站

18	湖北民族学院机构知识库	-	○	网站无法访问
19	华东师范大学机构知识库	-	○	网站无法访问
20	华南理工大学机构知识库	-	○	未找到网站
21	华南师范大学机构知识库	-	○	未找到网站
22	华中农业大学机构知识库	-	○	未找到网站
23	华中师范大学机构知识库	华中师范大学智汇云版权声明	●	
24	兰州大学机构知识库	-	◑	无相关内容
25	辽宁大学机构知识库	-	◑	无相关内容
26	南京师范大学机构知识库	-	○	未找到网站
27	南京医科大学机构知识库	-	○	网站无法访问
28	南京艺术学院机构知识库	-	○	网站无法访问
29	南开大学机构知识库	-	○	网站无法访问
30	内蒙古大学机构知识库	-	○	网站无法访问
31	清华大学学者库	清华学者库服务	●	
32	三明学院机构知识库	-	○	未找到网站
33	厦门大学学术典藏库	厦门大学学术典藏库（XMUIR）相关政策(试行)	●	

34	山东大学机构知识库	山东大学机构知识库政策	●	
35	山东师范大学机构知识库	-	○	网站无法访问
36	陕西师范大学机构知识库	-	○	未找到网站
37	上海交通大学机构知识库	-	○	未找到网站
38	沈阳师范大学机构知识库	-	○	网站无法访问
39	四川大学机构知识库	-	○	网站无法访问
40	同济大学机构知识库	同济大学机构知识库版权说明	●	
41	武昌首义学院机构知识库	-	○	网站无法访问
42	武汉大学机构知识库	武汉大学机构知识库开放获取相关说明	●	
43	武汉华夏理工学院机构知识库	-	○	未找到网站
44	西安电子科技大学机构知识库	-	○	未找到网站
45	西安交通大学机构知识库	西安交通大学机构知识门户开放获取政策	●	
46	西北工业大学机构知识库	-	○	网站无法访问
47	西南交通大学机构知识门户	-	○	未找到网站
48	浙江大学机构知识库	-	○	网站无法访问
49	中国海洋大学机构	-	○	网站无

	知识库			法访问
50	中国矿业大学机构知识库	-	○	网站无法访问
51	中国人民大学机构知识库	-	◑	无访问权限

注：●表示访问到了机构知识库网站，且获取到了相关内容；◑ 表示访问到了机构知识库网站，但没有获取到相关内容；○表示没有访问到机构知识库网站。

4.1.2 文件发布时间分析

在从数据中心和机构知识库获取到的 102 部科学数据管理相关政策文件中，共有 72 部文件标明了发布时间，占 70.6%，其中来自数据中心的有 69 部，来自机构知识库的有 3 部。对数据政策发布时间分布特征进行分析可知（见图 1）：1）获取到的数据政策文件最早可追溯至 2006 年的《海洋数据应用记录格式》，该记录获取自国家极地科学数据中心，该平台对相关标准规范按学科、分条目做了详细的整理著录；2）数据政策文件发布高度集中的时间区间在 2016-2020 年，并在 2019 年左右达到峰值，这与近年来从知识范式到数据范式的数据科学发展时间历程以及《科技部 财政部关于发布国家科技资源共享服务平台优化调整名单的通知》的发布时间基本吻合；3）结合图中所示与调研情况来看，数据政策的发布数量仍有一定的上升空间。如国家空间科学数据中心的数据政策栏目正在建设当中，国家天文科学数据中心拟制定一系列的天文科学相关数据标准规范等，这也从侧面反映出了科学数据管理活动的蓬勃发展之势。

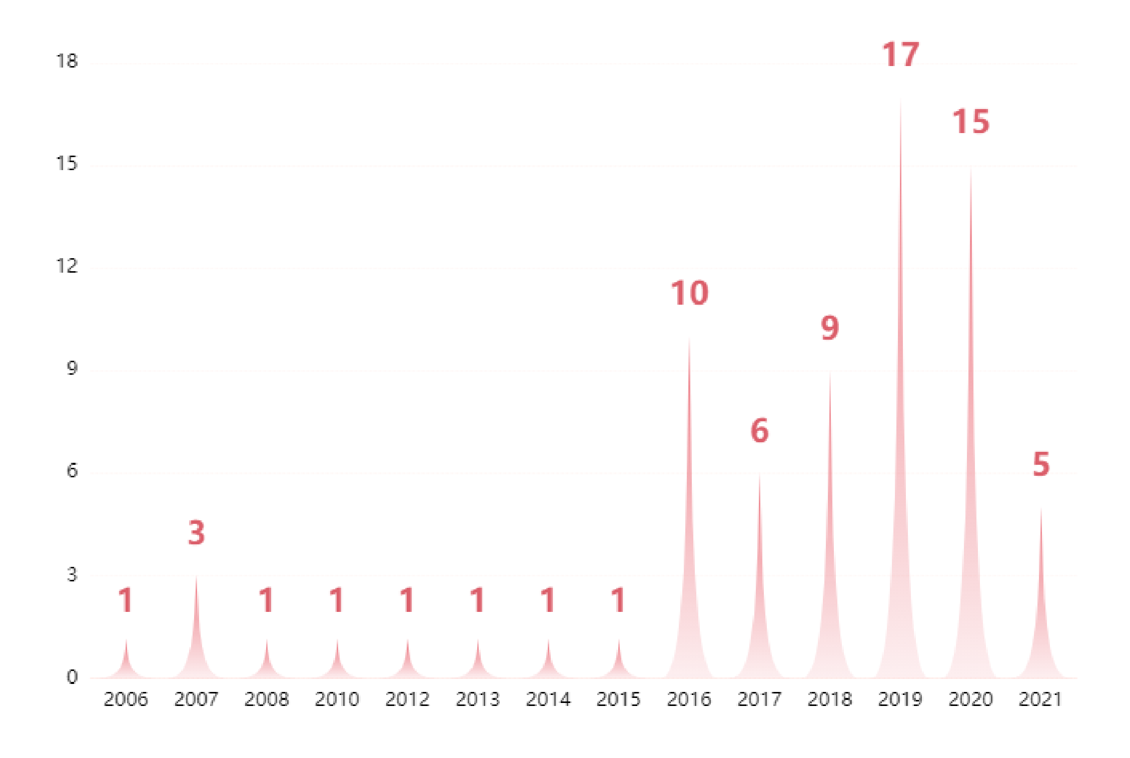
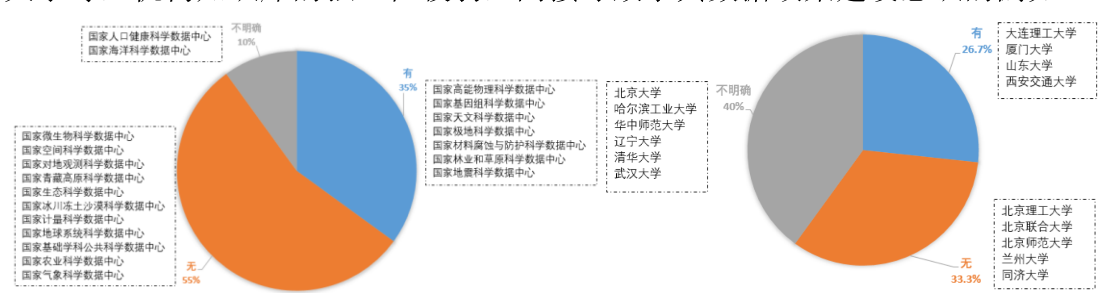


图 1 政策文件发布时间分布图

4.1.3 文件发布渠道分析

对 35 家能访问到的数据中心（20 家）和机构知识库（15 家）官方网站的科学数据管理政策发布渠道进行分析发现：1）20 家数据中心中，明确在网站首页设置数据政策或标准规范栏目并将其集中发布管理的有 7 家（占 35%），设置了集中发布管理相关数据文档专区但没有明确将其命名为政策栏目的有 2 家（占 10%），其余 11 家（占 55%）数据中心网站均无明确在网站首页专门设置数据政策栏目；2）15 家机构知识库中，专门设置了政策栏目的有 4 家（占 26.7%），没有明确设置政策栏目但有政策文档专区的有 6 家（40%），其余 5 家（占 33.3%）无相关栏目；3）从统计数据上看，机构知识库更倾向于不明确设置政策栏目而设置政策文档专区，常见的命名方式有“使用指南”“版权说明”“服务介绍”等。相比于数据中心的管理角色，机构知识库的服务性更强，且多数机构知识库是依附于高校图书馆建设，作为特色馆藏而存在，如北京科技大学、东南大学、中国海洋大学等，机构知识库的独立性较弱，间接导致了其数据政策建设意识的阙如。



（a）国家科学数据中心　　　　　　　（b）高校机构知识库

图 2 政策文件发布渠道分析图

4.1.4 文件来源渠道分析

对获取到全文的 75 部数据中心数据政策和 12 部机构知识库数据政策的来源渠道进行分析发现：1）网站发布的政策文件主要来源于三个方面：一是上级政策，即由科技部等政府部门颁布的相关法律法规、平台托管机构发布的各项管理办法，如《科学数据管理办法》《中科院科学数据管理与开放共享办法》等；二是领域标准，包括来自国内外的学科内部已有的通用标准或行业规范，如国际虚拟天文台联盟（The International Virtual Observatory Alliance, IVOA）规范系列等；三是平台自建，即由平台自身制定的该领域相关数据标准或本平台数据汇交流程控制办法，如《高能物理科学数据合作共享制度》《空间科学数据汇交指南》等。2）统计发现（图 3）：从数据中心网站获取到的 75 部数据政策中，有 13 部（17.3%）来源于上级政策，32 部（42.7%）来源于领域标准，30 部（40%）来源于平台自建；从机构知识库网站获取到的 12 部数据政策均属于平台自建。3）对比来看，数据中心数据管理政策的来源渠道更广、规范化程度更高、政策内容更为详细。各数据中心的数据政策建设基本遵循“援引+自建”的模式，即援引上级科学数据管理规定或国际领域通用标准的同时，积极制定各自领域的数据标准或建立本平台数据汇交流程控制办法。如国家天文科学数据中心主要援引了国际虚拟天文台联盟（IVOA）的各项数据标准，同时拟制定的标准规范包括《天文科学数据分类与编码》、《天文科学数据汇交规范》、《天文科学数据质量管理规范》等。

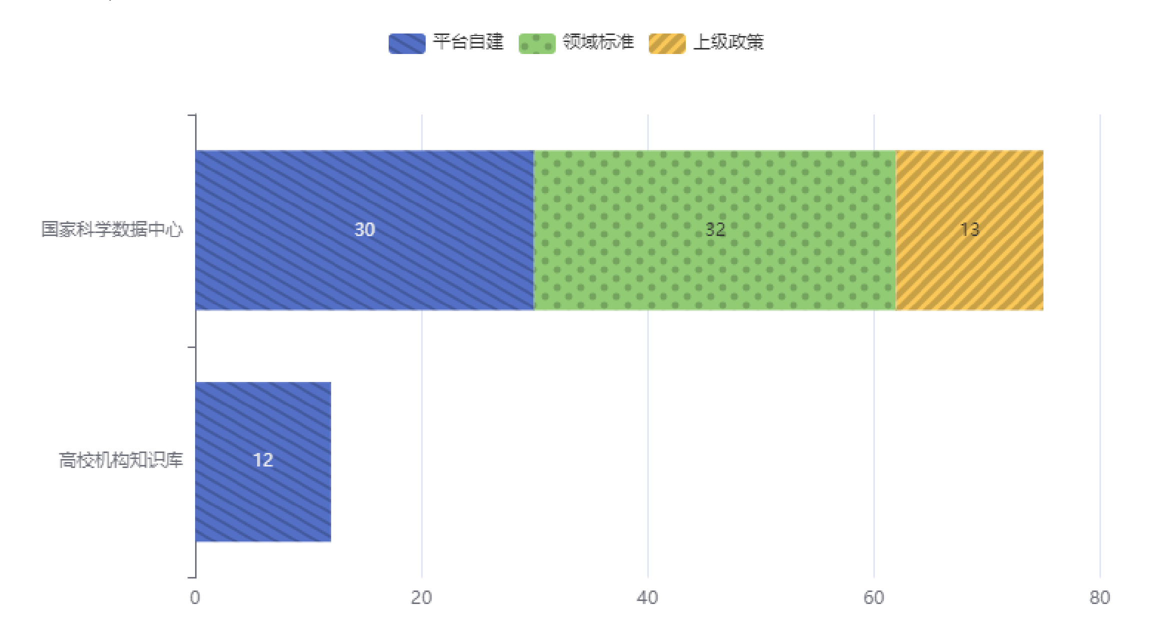


图 3 政策文件来源渠道统计图

4.1.5 政策内容热点分析

高频词可以反映出科学数据管理政策关注的焦点内容，因此我们深入到数据中心和机构知识库数据政策的内容单元，挖掘出其中的高频词（表 2 和表 3）并依据高频词表绘制词云图（图 4），通过两相对比发现：1）二者对“数据”本身的重视程度有所差别。数据中心将“数据”摆在绝对的领导地位，“数据”一次出现的频次断崖式远超其他词汇，而机构知识库相对来说则更加重视“成果”“作品”等更为宽泛的内容，“数据”一词的词频仅排在第四。2）从显性特征来看，数据中心的数据政策表现出了较强的学科领域特征，如“地震”“观测”“科技”等词汇出现频次较高，这也说明了不同领域的数据政策建设的完善程度不一，如国家地震科学数据中心的标准规范相对更加健全，内容更加丰富全面；而机构知识库的数据政策则更多地揭示了平台的功能定位特征，如“作者”“我校”“存储”“提交”等词汇占比较大，体现了机构知识库平台作为供本校师生使用的高校自有学术成果存储平台的核心价值。3）从关注重点上看，数据中心更加注重规范管理，呈现出“国家”“单位”“台站”“中心”“项目”等的分级管理制度，突出“管理”“服务”等内容，最终达到“科学数据共享”的目的；而机构知识库的数据政策更多地涉及开放获取相关内容，以“授权”“协议”“许可”等方式，将成果“提交者”、高校、“图书馆”等各方主体融合，打造以“保存”为主的，可供“免费”获取的知识交流平台。4）初步考察数据政策的 FAIR 特征发现，数据中心和机构知识库的数据政策均在一定程度上显露出了 FAIR 特征，但侧重点有所不同。数据中心更注重对数据的标识、描述与引用（F1、F2、I3），而机构知识库则把中心放在“协议”“许可”“授权”(A1.1、A1.2、R1.2）上，突出成果的可访问和可重用。详见表 3 表 4。

表 3 国家科学数据中心科学数据管理政策高频词汇表（TOP60）

序号	词语	词频	序号	词语	词频	序号	词语	词频
1	数据	6478	21	科技	488	41	服务	324
2	地震	1803	22	值域	488	42	记录	324
3	信息	1252	23	字符串	476	43	**描述**	322
4	**代码**	1040	24	**原始数据**	469	44	机构	314
5	名称	1002	25	版本号	461	45	台站	302
6	资源	993	26	内容	460	46	定义	300
7	科学数据	964	27	基础	446	47	管理	297

8	文本	864	28	项目	411	48	日期	296
9	**数据类型**	814	29	元素	407	49	科学数据共享	295
10	自由	806	30	负责	394	50	数据库	274
11	观测	789	31	实验	389	51	扩展	272
12	单位	692	32	实体	386	52	**标识**	263
13	分类	660	33	相关	380	53	共享	262
14	英文名称	648	34	..	376	54	类型	260
15	**标准**	638	35	时间	347	55	**引用**	252
16	次数	621	36	属性	344	56	发布	252
17	V1.0	537	37	用户	343	57	提供	242
18	汇交	520	38	条件	339	58	GB	240
19	**格式**	502	39	包括	334	59	研究	237
20	国家	489	40	选项	334	60	中心	235

注：加粗部分为 FAIR 相关词语。

表 4 高校机构知识库科学数据管理政策高频词汇表（TOP60）

序号	词语	词频	序号	词语	词频	序号	词语	词频
1	成果	233	21	目的	52	41	论文	27
2	作品	142	22	知识	44	42	图书馆	27
3	作者	117	23	权利	43	43	推荐	27
4	数据	112	24	**提交者**	43	44	门户	27
5	提交	93	25	信息	40	45	免费	26
6	存储	89	26	资源	39	46	传播	26
7	提供	84	27	格式	38	47	形式	26
8	内容	82	28	科研	36	48	PKU	25
9	机构知识库	78	29	方式	35	49	**访问**	25
10	我校	78	30	**许可**	34	50	保留	25
11	机构	77	31	系统	34	51	撤回	24
12	学者	75	32	**协议**	33	52	部门	24
13	IR	70	33	**长期**	32	53	XJTU	24
14	保存	67	34	情况	31	54	Academic	24
15	学术	66	35	相关	31	55	Hub	24

16	政策	64	36	**授权**	30	56	包括	23
17	全文	60	37	存档	30	57	永久	23
18	获取	57	38	开放	30	58	SDU	23
19	用户	56	39	文件格式	28	59	北京大学	22
20	版权	54	40	管理	27	60	本人	22

注：加粗部分为 FAIR 相关词语。



（a）国家科学数据中心　　　　　（b）高校机构知识库

图 4 高频词词云图（TOP200）

4.2 FAIR 原则视角下我国科学数据管理政策的对齐分析

4.2.1 FAIR 锚点词表

锚点词是用于定位 FAIR 原则文本与政策文本，具有特征明显、辨识度高、涵盖性强等特点的词语。FAIR 原则的原文是英文形式，目前国内学者对其进行的评介有几种不同的译文版本[50][51][52][53]，通过对其进行分析比对、凝聚提炼，我们根据同义关系为 FAIR 原则的四条总原则赋予锚点词，并根据相关关系从 FAIR 原则的十五条子原则中析出锚点词，从而得到了 FAIR 锚点词表（表 5）。需要说明的是：1）国家基因组科学数据中心和国家天文科学数据中心的相关政策文本为英文形式，所以我们直接从 FAIR 原则原文中析出锚点词，即表中加粗的部分；2）FAIR 子原则的 A1 和 R1 原则分别是对 A1.1、A1.2 和 R1.1、R1.2、R1.3 的概括性说明，因此没有特别析出锚点词来进行匹配，其对齐程度可由其包含的子原则反映出来。

表 5 FAIR 锚点词表

FAIR 原则	同义锚点词	FAIR 子原则	相关锚点词
Findable	可发现 可查找	F1:(Meta)data are assigned a globally **unique and persistent identifier**	标识符 唯一标识

	可检索 查询	F2:Data are described with rich **metadata** (defined by R1 below)	元数据
		F3:Metadata clearly and explicitly include the **identifier** of the data they describe	标识符
		F4:(Meta)data are registered or indexed in a **searchable** resource	检索
Accessible	可获取 可访问 可存取 存储 储存	A1:(Meta)data are retrievable by their identifier using a standardised communications protocol	-
		A1.1:The **protocol** is open, free, and universally implementable	协议
		A1.2: The protocol allows for an **authentication** and **authorisation** procedure, where necessary	身份验证 授权
		A2:Metadata are accessible, even when the data are no longer **available**	长期保存
Interoperable	可交互 可互操作 可融合 可整合 理解 阅读 识别	I1:(Meta)data use a formal, accessible, shared, and broadly applicable language for **knowledge representation**	知识表示 知识组织
		I2:(Meta)data use **vocabularies** that follow FAIR principles	术语系统 词表 本体
		I3:(Meta)data include qualified **references** to other (meta)data	引用
Reusable	可重用 再利用 重复使用 再用	R1: (Meta)data are richly described with a plurality of accurate and relevant attributes	-
		R1.1:(Meta)data are released with a	使用许可

再次使用	clear and accessible data **usage license**	
	R1.2:(Meta)data are associated with detailed **provenance**	数据来源
	R1.3:(Meta)data meet domain-relevant **community standards**	领域标准

4.2.2 FAIR 原则到政策文本对齐

如前所述，FAIR 原则到政策文本的对齐可从语句和语义两个层级，分别结合 FAIR 总原则和子原则，共四个方面展开，即 FAIR 总原则到政策文本的语句对齐及其语义对齐，FAIR 子原则到政策文本的语句对齐及其语义对齐。以 FAIR 总原则中的 Findable 为例，将“可发现”作为锚点词，在语句对齐阶段，共定位到 5 个相关语句；深入到语义层面分析发现，其中 4 个为真锚点相关语句，1 个为伪锚点相关语句（见表 6）。对其他锚点词的定位与筛选方法同理，需要说明的是，对于一些语义明确的锚点词如“元数据”“标识符”等，其作为专有名词内涵清晰且少有歧义，则一般不再进行语义级别的筛查。

表 6 FAIR 原则到政策文本对齐示例

FAIR 总原则	对应政策文本	锚点真伪性
Findable	科学数据应按照高能物理科学数据中心数据汇交要求的标准规范加工处理，确保汇交科学数据***可发现***、可获取、互操作和可重复利用	1
	科技计划项目所汇交的科学数据应按照相关科学数据汇交要求的标准规范加工处理，确保数据***可发现***、可获取、可互操作和可重复利用。	1
	科学数据应按照分等级、***可发现***、可访问、可重用的原则，适时向院内外用户开放共享	1
	可发现：每个数据集都应被赋予符合国家标准、唯一且长期 不变的标识符，并配有规范的元数据描述，易于发现和定位。	1
	投稿指南：通过机构成果的数据分析，***可发现***对机构比较友好的期刊清单，进而提升机构教职工投稿的录用率。	0

注:“1”为真锚点,“0”为伪锚点

4.2.3 对齐结果可视化分析

FAIR 原则到政策文本对齐的矩阵热力图描述了我国 18 家数据中心和 10 家机构知识库科学数据管理政策的 FAIR 化程度及其分布情况。如图 5、图 6 所示，每一个矩阵元素中的数字代表该机构政策文本匹配到对应 FAIR 原则锚点词的次数，颜色代表匹配强弱程度，次数越多，匹配程度越强，颜色越深。且为提高颜色映射效果，我们根据次数最高值将次数分段进一步划分为 5 个区域，如在 FAIR 总原则的对齐中，次数最高值为 45，那么 0-9 次则为低匹配程度区域，10-18 次为较低匹配程度区域，19-27 次为中匹配程度区域，28-36 次为较高匹配程度区域，37-45 次为高匹配程度区域，FAIR 子原则的对齐同理。这样可以使次数分段区域内颜色相近, 而区域间颜色差异显著，从而突出匹配程度，减少视觉杂乱。

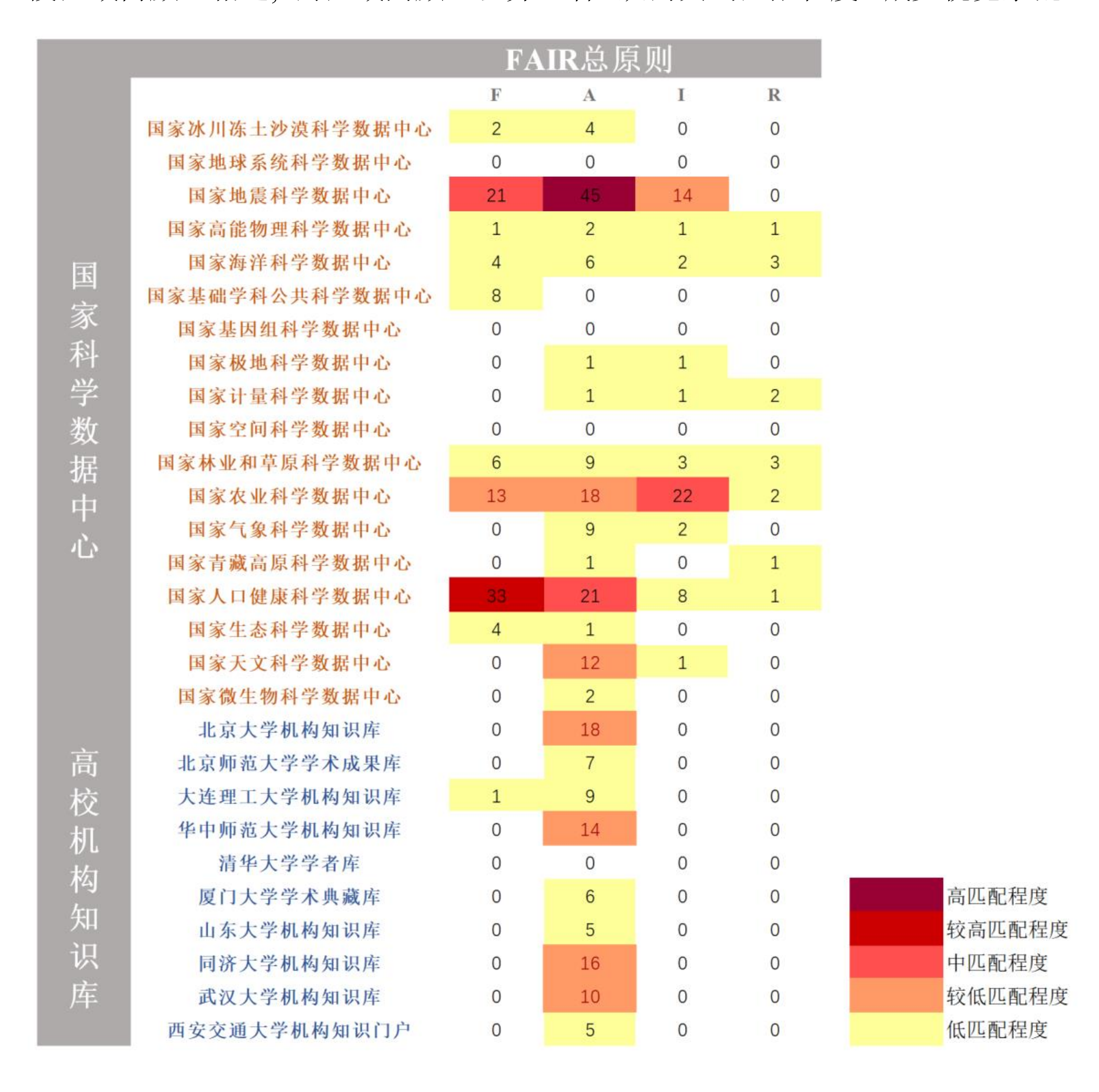
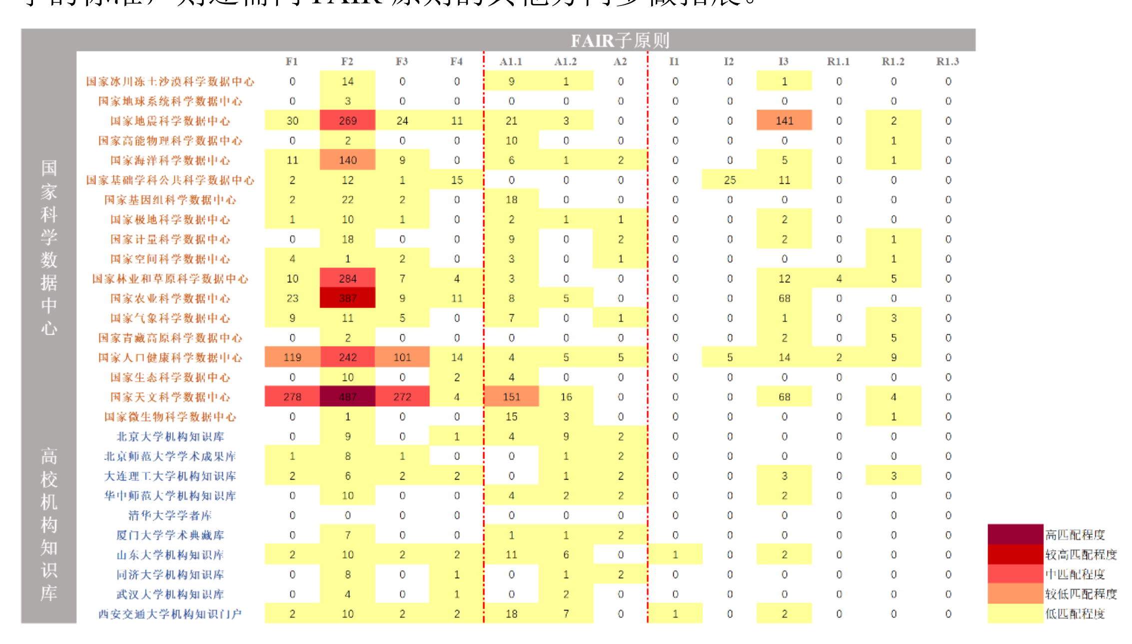


图 5 FAIR 总原则到政策文本对齐矩阵热力图

纵观 FAIR 原则到政策文本对齐的概况，不难发现：1）我国的科学数据管理政策中已经蕴含着一些 FIAR 基础理念，但是总体匹配程度偏低。总体来看，FAIR 的四条原则在政策文本中均有提及，并无遗漏现象，但是大多数仍处于无匹配、低匹配和较低匹配的状态，尚未形成系统化、组织化和显性化的 FAIR 思想。2）从 FAIR 原则的纵向匹配现状可以看出，政策本文中对 F 和 A 原则关注较多而对 I 和 R 原则涉及较少。可发现是实现数据 FAIR 化的基础[54]，而可获取是数据重用的前提条件之一，对这两者的关注是科学数据迈向 FAIR 化的必由之路，由此可见，我国的科学数据管理 FAIR 化尚处于起步阶段，政策方面还有待进一步深入完善。3）从机构的横向对比情况可以发现，数据中心政策文本对 FAIR 原则的整体匹配程度要优于机构知识库，但二者侧重有所不同。大部分数据中心的政策文本涉及到了两条及以上的 FAIR 原则，多数还是集中在基础的 F 和 A 原则上，少数如国家地震科学数据中心、国家农业科学数据中心已注意到了 I 原则的重要性，FAIR 化程度相对较为理想；而机构知识库一般只覆盖到了一或两条 FAIR 原则，且集中在 A 原则上，对其他原则尤其是 R 原则的关注不够，这种情况反映出机构知识库对开放获取运动较强的促进作用，但若要达到科学数据开放共享的标准，则还需向 FAIR 原则的其他方向多做拓展。



		FAIR子原则												
		F1	F2	F3	F4	A1.1	A1.2	A2	I1	I2	I3	R1.1	R1.2	R1.3
国家科学数据中心	国家冰川冻土沙漠科学数据中心	0	14	0	0	9	1	0	0	0	1	0	0	0
	国家地球系统科学数据中心	0	3	0	0	0	0	0	0	0	0	0	0	0
	国家地震科学数据中心	30	269	24	11	21	3	0	0	0	141	0	2	0
	国家高能物理科学数据中心	0	2	0	0	10	0	0	0	0	0	0	1	0
	国家海洋科学数据中心	11	140	9	0	6	1	2	0	0	5	0	1	0
	国家基础学科公共科学数据中心	2	12	1	15	0	0	0	0	25	11	0	0	0
	国家基因组科学数据中心	2	22	2	0	18	0	0	0	0	0	0	0	0
	国家极地科学数据中心	1	10	1	0	2	1	1	0	0	2	0	0	0
	国家计量科学数据中心	0	18	0	0	9	0	2	0	0	2	0	1	0
	国家空间科学数据中心	4	1	2	0	3	0	1	0	0	0	0	1	0
	国家林业和草原科学数据中心	10	284	7	4	3	0	0	0	0	12	4	5	0
	国家农业科学数据中心	23	387	9	11	8	5	0	0	0	68	0	0	0
	国家气象科学数据中心	9	11	5	0	7	0	1	0	0	1	0	3	0
	国家青藏高原科学数据中心	0	2	0	0	0	0	0	0	0	2	0	5	0
	国家人口健康科学数据中心	119	242	101	14	4	5	5	0	5	14	2	9	0
	国家生态科学数据中心	0	10	0	2	4	0	0	0	0	0	0	0	0
	国家天文科学数据中心	278	487	272	4	151	16	0	0	0	68	0	4	0
	国家微生物科学数据中心	0	1	0	0	15	3	0	0	0	0	0	1	0
高校机构知识库	北京大学机构知识库	0	9	0	1	4	9	2	0	0	0	0	0	0
	北京师范大学学术成果库	1	8	1	0	0	1	2	0	0	0	0	0	0
	大连理工大学机构知识库	2	6	2	2	0	1	2	0	0	3	0	3	0
	华中师范大学机构知识库	0	10	0	0	4	2	2	0	0	2	0	0	0
	清华大学学者库	0	0	0	0	0	0	0	0	0	0	0	0	0
	厦门大学学术典藏库	0	7	0	0	1	1	2	0	0	0	0	0	0
	山东大学机构知识库	2	10	2	2	11	6	0	1	0	2	0	0	0
	同济大学机构知识库	0	8	0	1	0	1	2	0	0	0	0	0	0
	武汉大学机构知识库	0	4	0	1	0	2	0	0	0	0	0	0	0
	西安交通大学机构知识门户	2	10	2	2	18	7	0	1	0	2	0	0	0

高匹配程度
较高匹配程度
中匹配程度
较低匹配程度
低匹配程度

图 6 FAIR 子原则到政策文本对齐矩阵热力图

深入到对 FAIR 子原则到政策文本对齐矩阵热力图的观察发现，我国科学数据管理政策对 FAIR 原则的关注程度呈阶梯状分布，且显著地分为三个层次：

1）对 F 原则高度关注，形成以“高原山地”为主的第一阶梯。F 原则主要包含了三个方面内容：元数据、标识符以及可检索。政策文本对 F 原则的关注集中

在 F2 原则上，即用丰富的元数据来描述数据。元数据是关于数据的数据，对数据进行管理鲜有不提及元数据的。对元数据的规定越详细，则对数据的管理越规范、越便利，因此 FAIR 子原则到政策文本匹配次数的“山地”最高峰也出现在 F2 上；其次是对标识符的规定，F1 原则要求为数据赋予全局唯一且持久的标识符，F3 原则要求标识符包含在元数据中，均是围绕标识符对数据做出要求，因此政策文本对这两条原则的匹配情况较为一致；最后是 F4 原则要求数据在可检索的资源中进行注册，从图上看，政策文本对该原则的匹配程度不如 F 原则的其他三条，但是需要指出的是，国家科学数据中心和高校机构知识库本身即是用于注册、提交和检索数据的平台。因此无论从文本本身还是实践情况来说，F 原则在我国的科学数据管理活动中都是有广泛基础的。

2）对 A 原则较为关注，形成以“平原丘陵”为主的第二阶梯。A 原则从数据的通信交流（协议、身份验证）和长期保存两方面做出规定。从图 6 中可以看出，政策文本对 A 原则的对齐分布比较平均，但是大部分停留在低程度的匹配上。相比较而言，数据中心与机构知识库在数据的通信交流方面形成了各自的“丘陵地带”，数据中心侧重于通过协议交流数据，而机构知识库强调对用户的身份验证与作者授权；在数据的长期保存方面，机构知识库的对齐情况要略优于数据中心。形成这种差异的主要原因在于二者的服务对象和功能定位不同，数据中心旨在面向广大科研人员提供数据服务，采用协议的方式进行数据交流更为便捷有效，而机构知识库更倾向于助力本校科研成果的开放获取，因此更重视数据成果知识产权的认证与保护，以及对成果的长期保存。

3）对 I 原则和 R 原则缺乏关注，形成了以“丘陵盆地”为主的第三阶梯。I 原则通过对数据的形式表示和限定引用的控制来促进数据交互，R 原则描述了数据重用所需要的一系列相关属性，包括使用许可、数据来源、领域标准等。从图上可以看出，政策文本对 I 和 R 原则的对齐状况不甚乐观，呈现出大部分无匹配的“盆地”地区，尤其是对数据的形式表示、使用许可、领域标准几个方面很少涉及；但是部分机构政策对数据的限定引用和数据来源有所关注，形成了为数不多的“丘陵”地带。限定引用是数据交互的常见形式，明确数据来源是数据重用的内在要求，由此可见，我国数据政策在可互操作和可重用方面的关注仍停留在较浅层面，而对深层次的数据组织方面的内容，如数据本身的表示形式、背后所使用的领域标准等，则还有进一步提升的空间。

结合 FAIR 总原则和子原则到政策文本对齐矩阵热力图来看，可以发现我国

科学数据管理政策的 FAIR 程度整体表现出一种“不均衡”的状态，这种“不均衡”状态表现为三个方面：

一是政策文本包含的 FAIR 理念与实际的 FAIR 表述之间存在差距。从图上可以看出，FAIR 总原则锚点词对政策文本的锚定效果相比于子原则锚点词要差，在机构知识库向 F 原则的对齐上可见一斑。这种情况说明我国的科学数据管理政策中已经孕育出一定的 FAIR 理念，只是在具体的文字表述上尚未与 FAIR 原则接轨，需要进一步加强对 FAIR 原则的理解与实施[55]。

二是政策文本在四条原则上表现出的匹配程度不均。这一点 FAIR 总原则与子原则的调查结果一致，我国科学数据管理政策在向 FAIR 四条原则的对齐程度上呈现出依次递减的趋势。这种趋势既符合 FAIR 各条原则本身逐渐深入最终达到可重用目的的逻辑架构，也反映出了我国科学数据管理与共享的 FAIR 化程度正处于起步阶段的现实情况。

三是机构内部之间的 FAIR 化程度不一，包括数据中心与机构知识库之间，各数据中心之间以及各机构知识库之间的不均衡。相对来说，数据中心整体的数据政策 FAIR 化程度要优于机构知识库，但同时数据中心内部各机构的 FAIR 化程度差异要大于机构知识库。国家科学数据中心本身是为了推动科学数据等科技资源的开放共享，该主旨与 FIAR 原则的理念更为接近，但是不同领域的数据中心依托于各自行业的研究机构而存在，数据政策建设进度不一，这将形成阻碍数据交互的数据孤岛，为交叉学科的研究带来不利影响；而机构知识库借助中国高校机构知识库联盟这一行业组织，遵守统一的相关政策、平台系统、标准规范等，有明确的建设方向与目标，总体 FAIR 程度也较为趋同。因此，数据中心或可借鉴这种模式，积极推动数据政策的统一建设，提高整体数据 FAIR 化程度，促进各中心、各行业、各领域之间的数据交流共享。

5 结语

本文在全方位、多角度了解我国科学数据管理政策内外部特征的基础上，提出了一种基于“锚点词”的 FAIR 原则到政策文本的对齐方法，综合运用特征统计、可视化分析等手段，得到以下主要结论：

从总体对齐结果来看，我国科学数据管理政策中已经蕴含着一些 FIAR 基础理念，但是总体匹配程度偏低；综合每条 FAIR 细则的匹配程度来看，我国科学数据管理政策对 FAIR 原则的关注程度呈阶梯状分布，且显著地分为三个层次，即：对 F 原则高度关注，对 A 原则较为关注，对 I 原则和 R 原则缺乏关注；纵

观我国科学数据管理政策的 FAIR 程度，发现其整体表现出一种“不均衡”的状态，表现在三个方面：一是政策文本包含的 FAIR 理念与实际的 FAIR 表述之间存在差距，二是其次是政策文本在四条原则上表现出的匹配程度不均，三是最后是机构内部之间的 FAIR 化程度不一。

同时，本研究也具有一定局限性。一是在研究过程中，笔者虽不拘泥于平台网站，并试图从其他来源获取政策文件全文，但最终无法获取全文的情况仍无可避免，这虽然一定程度上是由于平台建设存在短板，但也使得本文研究结果与真实情况或许存在一定偏差，未能完全还原我国科学数据管理政策的 FAIR 化面貌；二是利用 FAIR 锚点词表进行字符串匹配的方式只能探测出语义较为明确的、表述相对成熟的 FAIR 语句，而无法对某些语义相同而表述各异的 FAIR 锚点进行抓取、归纳与分析，这为我们对政策文本 FAIR 化特征与程度的研判造成了一定障碍。这些将是未来值得深入研究和继续探索的重点。

参考文献：

[1] Open Access Week. 2021 Open Access Week Theme to be “It Matters How We Open Knowledge: Building Structural Equity” [EB/OL].[2021-10-10]. https://www.openaccessweek.org/profiles/blogs/2021-theme-announcement-english

[2] H. van Vlijmen, A. Mons, A. Waalkens, W. Franke, A. Baak, G. Ruiter, ... & J.-M. Neefs. The need of Industry to go FAIR. Data Intelligence Special Issue 1: Emerging FAIR Practices (2020), 276–284. doi: 10.1162/dint_a_00050

[3] FORCE11.GUIDING PRINCIPLES FOR FINDABLE, ACCESSIBLE, INTEROPERABLE AND RE-USABLE DATA PUBLISHING VERSION B1.0 [EB/OL].[2021-10-10].https://www.force11.org/fairprinciples.

[4] 中国科学院办公厅. 中国科学院印发《中国科学院科学数据管理与开放共享办法（试行）》[EB/OL].[2021-10-20]. http://www.go.cas.cn/gzdt/201902/t20190221_4679909.html

[5] 中华人民共和国科学技术部.《新一代人工智能伦理规范》发布[EB/OL].[2021-09-27]. http://www.most.gov.cn/kjbgz/202109/t20210926_177063.html

[6] 孔丽华,习妍,张晓林.数据出版的趋势、机制与挑战[J].中国科学基金,2019,33

(03):237-245.
[7] 张文萍,宋秀芬,魏银珍,李立睿.基于FAIR标准的科学数据融合体系研究[J].中国图书馆学报,2020,46(06):41-54.
[8] 秦顺.面向一流高校建设的图书馆科研数据管理服务研究——以整合 DLC-SH为视角[J].图书情报工作,2021,65(04):28-39.
[9] 刘建平. 科学数据用户相关性判断模型研究[D].中国农业科学院,2020.
[10]顾立平.科学数据权益分析的基本框架[J].图书情报知识,2014,157(1):34-51.
[11]邹中才,顾立平,张闪闪,等. 图书馆实施分阶段科研数据管理服务中的权益问题初探[J]. 情报理论与实践, 2016, 39(1): 64-69.
[12]朱大丽,介风. 从科学数据素养角度剖析国内外高校科学数据管理[J]. 图书馆学研究, 2015, 363(16): 53-59.
[13]蒋星. 国外高校图书馆科学数据管理调研现状与启示[J]. 图书馆界, 2020, 174(1): 32-37.
[14]张计龙,朱勤,殷沈琴. 美国社会科学数据的共享与服务[J]. 大学图书馆学报, 2013, 31(5): 13-17.
[15]Christine l.borgman,青秀玲. 科研数据共享的挑战[J]. 现代图书情报技术, 2013, 233(5): 1-20.
[16]邢文明. 我国科研数据管理与共享政策保障研究[D]. 武汉大学, 2014.
[17]邢文明,洪芳林,李晓妍.科学数据管理体系的二维视角——《科学数据管理办法》解读[J].图书情报工作,2019,63(23):30-37.
[18]秦顺,邢文明.开放·共享·安全:我国科学数据共享进入新时代——对《科学数据管理办法》的解读[J].图书馆,2019(06):36-42.
[19]邢文明,洪程.开放为常态,不开放为例外——解读《科学数据管理办法》中的科学数据共享与利用[J].图书馆论坛,2019,39(01):117-124.
[20]高瑜蔚,石蕾,朱艳华,胡良霖,顾立平.《科学数据管理办法》实施细则比较研究——以正式发布的11份细则为例[J].中国科技资源导刊,2019,51(03):1-10+17.
[21]白锐,吕跃.基于修正多源流模型视角的政策议程分析——以《科学数据管理办法》为例[J].图书馆理论与实践,2019(10):50-55.
[22]温亮明,李洋,余波.基于区块链技术的《科学数据管理办法》落实路径探析[J].现代情报,2021,41(08):136-146.
[23]李洋,温亮明.《科学数据管理办法》落实现状、影响因素及推进策略研究[J].

图书情报工作,2021,65(02):65-74.
[24]邢文明,汤雅静,秦顺．国外教育机构科研数据管理政策大纲解读及启示[J]．数字图书馆论坛，2019，180(5)：9-16.
[25]张瑶,顾立平,杨云秀,等．国外科研资助机构数据政策的调研与分析——以英美研究理事会为例[J]．图书情报工作，2015，59(6)：53-60.
[26]王丹丹,董金金,杨嘉敏．《科研数据管理国际联盟实用指南》研究及启示[J]．数字图书馆论坛，2021，203(4)：17-24.
[27]邢文明,杨玲．中美科学数据政策比较——以《科学数据管理办法》和《促进联邦资助科研成果获取的备忘录》为例[J]．图书馆论坛，：1-10.
[28]谢春枝,燕今伟．国内外高校科学数据管理和机制建设研究[J]．图书情报工作，2013，57(6)：12-17，38.
[29]赵小兰,刘桂锋．香港高校科研数据管理服务调查分析[J]．数字图书馆论坛，2018，169(6)：37-44.
[30]孟祥保．英国高校图书馆科研服务现状调研及启示[J]．图书情报工作，2017，61(13)：53-61.
[31]刘莉．基于机构知识库的科研数据管理分析[D]．山东理工大学，2020.
[32]陈大庆.英国科研资助机构的数据管理与共享政策调查及启示[J]．图书情报工作，2013，57(8)：5-11.
[33]王炼．美国联邦政府科学数据管理政策及实践[J]．全球科技经济瞭望，2018，33(7)：47-51.
[34]赵辉．美国地质调查局数据管理机制及启示[J]．数字图书馆论坛，2019，182(7)：24-30.
[35]刘凤红,彭琳．FAIR 原则背景下国际出版集团的数据政策和实践[J]．中国科技期刊研究，2021，32(2)：173-179.
[36]吴蓉,顾立平,刘晶晶．国外学术期刊数据政策的调研与分析[J]．图书情报工作，2015，59(7)：99-105.
[37]江洪,刘敬仪．国外期刊科学数据管理调查与分析[J]．图书情报工作，2019，63(9)：127-134.
[38]李芳薇,程瑾,张群,等．国外图书馆生物医学科研数据管理服务及启示[J]．中华医学图书情报杂志，2015，24(8)：5-10.
[39]李芳薇,王璇,杨志滨,等．基于生物医学科研数据仓储的图书馆数据管理服务

[J]. 中华医学图书情报杂志, 2019, 28(3): 1-5.
[40]王卷乐,王明明,石蕾,等. 科学数据管理态势及其对我国地球科学领域的启示[J]. 地球科学进展, 2019, 34(3): 306-315.
[41]王淑强,王卷乐,李扬,等. 基于文献计量学的国际地球科学数据管理研究进展[J]. 全球变化数据学报(中英文), 2020, 4(3): 299-313.
[42]陈廉芳. 医学科学数据馆员制度建设[J]. 图书馆学研究, 2018, 431(12): 92-96.
[43]周玉琴,邢文明.我国科研数据管理与共享政策体系研究[J].中华医学图书情报杂志,2018,27(08):1-7.
[44]薛秋红,徐慧芳.西方国家科研机构科学数据管理政策要素研究[J].情报理论与实践,2021,44(07):191-196+124.
[45]中华人民共和国科学技术部.科技部 财政部关于发布国家科技资源共享服务平台优化调整名单的通知[EB/OL].[2021-09-08].http://www.most.gov.cn/xxgk/xinxifenlei/fdzdgknr/qtwj/qtwj2019/201906/t20190610_147031.html
[46]中国高校机构知识库联盟.中国高校机构知识库联盟章程[EB/OL].[2021-09-08]. http://chair.calis.edu.cn/pages/list.html?id=e422949c-ef68-4baa-b265-28b124df86ce
[47]涂志芳,杨志萍. 我国科学数据管理与共享实践进展:聚焦两种主要模式[J]. 图书情报知识, 2021, 199(1): 103-112.
[48]Wilkinson, M. D. et al. A design framework and exemplar metrics for FAIRness. Sci. Data 5:180118 doi: 10.1038/sdata.2018.118 (2018).
[49]韦向峰,袁毅,张全,池毓焕.富媒体环境下语音和文本内容的对齐研究[J].情报工程,2019,5(02):17-27.
[50]段青玉,王晓光.人文社科数据出版平台 FAIR 原则应用调查研究[J].科技与出版,2019(04):6-11.
[51]邢文明,郭安琪,秦顺,杨玲.科学数据管理与共享的 FAIR 原则——背景、内容与实施[J].信息资源管理学报,2021,11(02):60-68+84.
[52]邱春艳.开放科学愿景下欧盟推进 FAIR 原则的路径、经验及启示[J].情报理论与实践,2021,44(05):199-205.
[53]宋佳,温亮明,李洋.科学数据共享 FAIR 原则：背景、内容及实践[J].情报资料工作,2021,42(01):57-68.

[54] Nick Juty, Sarala M, Stian Soiland, et al. Unique, Persistent, Resolvable: Identifiers as the Foundation of FAIR[J]. Data Intelligence, 2020, 2(1-2):30-39.

[55] A. Jacobsen, R. de Miranda Azevedo, N. Juty, D. Batista, S. Coles, R. Cornet, ... & E. Schultes. FAIR principles: Interpretations and implementation considerations. Data Intelligence Special Issue 1: Emerging FAIR Practices (2020), 11–29. doi: 10.1162/dint_r_00024